Transportation Research Part D 111 (2022) 103442

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Using explainable machine learning to understand how urban form shapes sustainable mobility

Felix Wagner [a,b,*], Nikola Milojevic-Dupont [a,b], Lukas Franken [a,c,e], Aicha Zekar [a,d], Ben Thies [a], Nicolas Koch [a], Felix Creutzig [a,b]

[a] Mercator Research Institute on Global Commons and Climate Change, Berlin, Germany
[b] Technical University Berlin, Berlin, Germany
[c] Institute for Theoretical Physics, University of Cologne, Cologne, Germany
[d] New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[e] School of Engineering, The University of Edinburgh, Edinburgh, Scotland, United Kingdom

ARTICLE INFO

Keywords:
Sustainable mobility
Urban form effects
Urban planning
Explainable machine learning

ABSTRACT

Municipalities are increasingly acknowledging the importance of urban form interventions that can reduce intra-city car travel in achieving more sustainable cities. Current academic knowledge for supporting such policies falls short in providing the spatial details required to plan specific interventions. Here, we develop an explainable machine learning framework to identify location-specific relevance of built environment for urban motorised travel, using a sample of 3.5 million car commutes over one year in Berlin and high-resolution urban form data. Results demonstrate that subcenters play a vital role in reducing commuting-related travel distance, giving support to the *15-minute city* hypothesis. Observed threshold effects of induced CO_2 emissions require low-carbon-policies targeted towards densifying the inner city while releasing peripheral low income communities from car dependence. This research provides a starting point for increasingly rich big data analyses of urban form for creating low-carbon and inclusive urban planning strategies.

1. Introduction

Global sustainability requires an urban science that clearly recognises and explicitly includes variation in urban conditions, and that renders fields such as engineering and artificial intelligence comprehensible for urban policy makers (Acuto et al., 2018). This becomes crystallised in the challenge of urban mobility, responsible for about 3 Gt CO_2-eq per year (Creutzig et al., 2016), but also causing air pollution and noise, reducing objective and subjective safety and hindering healthy active mobility (Seto et al., 2014; Woodcock et al., 2009; Creutzig and He, 2009; Watts et al., 2015). Urban planning and its role in modifying urban form and associated patterns of mobility and lifestyle has been identified as a main lever to address the environmental and social costs of urban transport. We observe a gap in the current literature as solutions remain conceptual, subject to aggregate statistics or limited to particular instances (Seto et al., 2014; Newman and Kenworthy, 1999). Here, we posit that available big data on both urban infrastructure and mobility behavior enables for the first time a nuanced and spatially fine-grained understanding of how urban form shapes sustainable or unsustainable urban mobility. Our contributions are three-fold. We first demonstrate how predictive urban form features are for predicting car-driving distance (VKT), both at the trip origin and destination. Then, we highlight the

* Corresponding author at: Technical University Berlin, Berlin, Germany.
E-mail address: wagner@mcc-berlin.net (F. Wagner).

https://doi.org/10.1016/j.trd.2022.103442
Received 28 March 2022; Received in revised form 18 August 2022; Accepted 22 August 2022
Available online 15 September 2022

relative importance of each independent urban form feature in the prediction. Finally, we demonstrate how the effect of a single urban form feature on VKT unfolds spatially and how it interacts with the given socio-demographic context.

2. Literature review

The scientific literature identified several solid lessons on urban form and induced mobility patterns. Three decades of empirical research show that compact development tends to yield shorter vehicle kilometer traveled (Newman and Kenworthy, 1999; Sims et al., 2014; Creutzig, 2016; Creutzig et al., 2015), while improving additional sustainability measures, such as increased safety or social capital (Ewing and Cervero, 2017; Ewing et al., 2018; Ding et al., 2018).

Compact development is typically analyzed through a set of features referred to as the *6D's of compact development*, which describe a location's destination access, density, distance to transit, diversity, design and demographic properties (Ewing and Cervero, 2010; Næss et al., 2017). Rationales to include the D variables are that they affect the utility of travel choices and thereby trip lengths (Ewing et al., 2015). For example, points of interest are located closer together due to higher density or land use diversity and are easier to reach due to higher destination accessibility (Stevens, 2017). Pedestrian friendly street design and shorter distances to transit make alternative modes more attractive (Stevens, 2017), but are less clear of an indicator for car VKT. Yet, they might have an interactive effect on VKT with other D-variables or could be relevant at particular locations (Ewing et al., 2015). Demographics are included to control for socio-economic differences that can affect travel choices of individuals due to, for example, varying financial constraints (Hong et al., 2014). While a number of studies have adopted the D variable framing, there has also been criticism on their general impact on travel (Stevens, 2017) and that the influence of all D variables is dependent on their relative location to the main center of the city, which many studies fail to address (Næss et al., 2021; Naess, 2011).

Most previous studies conclude that "destination accessibility", described by variables such as the "distance to the city center" (and thereby job access), "distance to a subcenter" or street connectivity, has the largest impact on vehicle kilometer traveled in comparison to other 6D's (Ding et al., 2018; Hong et al., 2014; Næss et al., 2017; Ewing and Cervero, 2017). "Socio-demographic" descriptors, such as income or social status, have been recognised to have a significant effect across continents, similar to "density" of population, jobs or roads (Hanson and Hanson, 1981; Ewing et al., 2015; Ding et al., 2018; Næss et al., 2017; Hong et al., 2014). "Diversity" of land use, such as mixed commercial and residential areas, impact vehicle kilometer traveled in particular for work-related trips (Ding et al., 2014; Næss et al., 2017). Other features, such as "design" of urban areas, e.g. sidewalk length or parking spaces, or "distance to transit" tend to have smaller although sometimes diverging impact on travel distance (Ewing et al., 2018; Silva et al., 2017; Ding et al., 2014; Næss et al., 2017).

Next to statistical approaches, previous work mainly on Europe (Næss et al., 2021, 2017; van de Coevering et al., 2021) and China (Zhao and Bai, 2019; Wang and Lin, 2019) has advanced the causal analysis of urban form effects on travel by combining quantitative and qualitative or utilising cross-sectional and quasi-longitudinal study designs that sufficiently control for residential self-selection effects. Nonetheless, previous research mainly relies on survey-based data, that is costly to collect, and may not depict the entire area under analysis with high spatial and temporal resolution (Liu et al., 2020). Hence, it provides little distinct quantitative guidance of context-specific urban form impact at street level (Silva et al., 2017). The literature still makes little use of novel available big data methods to compute city-wide effects at high spatial resolution. These methods hold promise to understand the specific conditions of potential urban form interventions towards novel sustainable urban development concepts, such as *15-min cities* (Moreno et al., 2021) – an urban planning idea that focuses on providing citizens access to basic amenities, such as work, education or healthcare within distances easily accessible in less than 15 min by foot or by bike (Moreno et al., 2021; Weng et al., 2019; Capasso Da Silva et al., 2020). A concept that aligns the focus on transit oriented urban density and resulting car independence highlighted by Newman and Kenworthy (1989, 1999, 2015) with the human-centric lens of Gehl (2010) and Jacobs (1961) to create sustainable, safe, and livable places.

Here, we demonstrate how machine learning and high-resolution data can identify location-specific relevance of the built environment for urban motorised travel. Our results support the relevance of subcenter design, and in particular of the *15-min city* (Moreno et al., 2021). We also find that commute related CO_2 emissions increase heavily towards the outer areas in the south-east and south-west of Berlin and that location of low-income households in these areas requires particular attention for urban planning.

3. Methods

We train two supervised machine learning models to predict average VKT within hexagon grid cells in the city of Berlin, using 18 features representing the 6D's of compact development. Fig. 1 provides an overview of the modeling framework. To construct our target variable and the predictive features representing the 6D's, we use seven datasets and perform extensive data cleaning before feature engineering. From a commercial mobility dataset, we calculate the average commuting distances within a hexagon grid cell. We complement this with commercial income data and open-access urban form, population, and social status data, to calculate the model features per grid cell.

We train two gradient boosting decision tree (GBDT) models with the calculated features as inputs and the average trip distance at the origin and at the destination as target. We apply feature selection and validate the prediction using a 5-fold spatial cross-validation procedure. Our results are compared with a median baseline model. Furthermore, we apply individual feature importance analysis based on SHAP values to derive a thorough understanding for the impact of individual features on the model prediction.

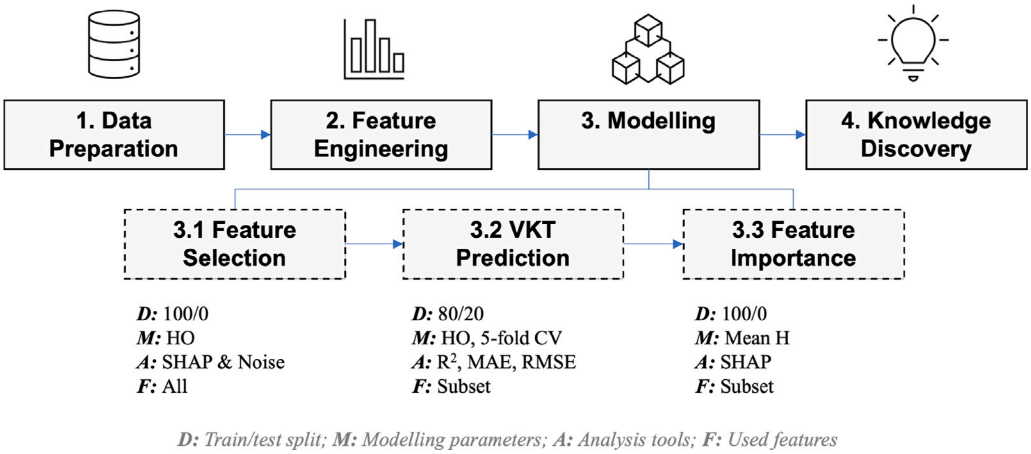

Fig. 1. Modeling framework. The four step modeling framework includes: **1.** Data Preparation, **2.** Feature Engineering, **3.** Modeling and **4.** Knowledge Discovery. The third step, modeling, is further separated into **3.1** Feature Selection, **3.2** VKT Prediction and **3.3** Feature Importance. Steps two to four are conducted twice to predict vehicle kilometers traveled by car for commute (VKT) using features at the trip origin and destination. For each step, we summarise information about the set-up chosen. ***D*** corresponds to how the train-test split was use. ***M*** corresponds to how the modeling parameters were adjusted i.e. via hyperparameter optimisation (HO), five-fold spatial cross-validation (*5-fold CV*) and average hyperparameter tuning from 5-fold CV (H). ***A*** indicates how the results were calculated and analyzed using SHAP values and a noise feature ($SHAP$ & $Noise$), the goodness of fit metrics (R^2, MAE, $RMSE$) or only SHAP values ($SHAP$). ***F*** indicates which features were used, either all 18 features (ALL) or a subset determined in the feature selection process ($Subset$).

With our modeling framework, we aim to close several methodological limitations in the existing literature, which have so far hindered the analysis of city wide effects with high spatial resolution. Previous literature has mostly relied on small sample sizes, not covering the area under analysis with high spatial and temporal resolution (Liu et al., 2020; Silva et al., 2017). A quasi-exclusive focus on the trip origin Boarnet et al. (2020), Choi and Paterson (2019), Wu et al. (2019), Tao et al. (2020), Salon (2015), Ding et al. (2018), Næss et al. (2017) and only very few studies incorporating the trip destination, such as (Maat and Timmermans, 2009) or (Liu et al., 2020), also limits the explanatory power of where effects occur. Similar to Maat and Timmermans (2009), we hypothesise that also the trip destination influences travel choices, as for example commuter trips in the morning hours might be drawn towards places where lots of jobs are located. In addition, only few studies analyze how the effect of individual features varies over their distribution (Ding et al., 2018), and no studies could be found that map these isolated effects in urban space. Further methodological shortcomings that were already highlighted in previous works (Ding et al., 2018; Hong et al., 2014; Ewing and Cervero, 2017; Næss et al., 2017) include analyzing single urban form features without controlling for others or not controlling for inter-feature correlation, spatial autocorrelation or residential self-selection effects. As a consequence, certain results might be biased, based on unlikely data instances or over-optimistic.

3.1. Data

3.1.1. Datasets

All datasets used in this study are collected for the area of Berlin (for more information about the study area, refer to SI, “1. Methods: Study Area”). To generate the target variable, VKT, the dataset used in this study was obtained from the commercial data provider INRIX. The dataset includes 34,208,544 unique trips originating, ending or crossing the city of Berlin. It contains waypoint data derived from the GPS signals of navigation systems of various sources, including connected cars, floating cars and commercial car fleets (such as taxis or freight trucks). The data was collected from 01.01.2017 until 31.12.2017. For urban form dimensions, we use street network data from (OpenStreetMap, 2021), land use data from (Copernicus Land Monitoring Service, 2018) and a LiDAR-generated streetspace dataset from (Berlin, 2014), where transit stations as well as parking areas are indicated respectively as points and polygons.

To capture socio-demographic effects, we utilise different socio-economic data sources. Recent studies have highlighted the importance of controlling for residential self-selection by considering socio-economic as well as residential preferences and travel attitudes as input variables (Næss et al., 2019a). However, due to a lack of openly available data at high resolution on preferences and attitudes, we cannot account for the part of residential self-selection that results from attitude-based residential preferences. We can only control for the aspects of residential self-selection effects, which are accounted for by socio-demographic variables, similar to Ding et al. (2018). We use gridded population density information at 100-m resolution, published as open-source information in Zensus (2011). We also include open government data on social status available at the level of the different planning districts of Berlin. The index is based on three main indicators, including unemployment-rate, child-poverty and transfer-payments, as well as three contextual indicators, including groups of severe poverty, immigration and housing conditions. It is defined as a static index representing the mean across all indicators per planning district at the latest measurement date, 31.12.2018, categorised on a four-level scale (very low, low, medium, high), and as a dynamic index, measuring the change (negative, unaltered, positive)

Table 1
Predictive features used to represent the 6D's of compact development. Each dimension is represented by one or several features, described in the table. An additional "noise" feature enables to monitor if the model overfits to random signal and to assess how the features compare to noise.

6D's of compact development	Feature in this study	Feature description
Destination accessibility	Distance to city center	Distance from hex. center to global center
	Distance to local center	Distance from hex. center to local center
	Street connectivity 1	Average edge number per node within hex.
	Street connectivity 2	Total number of edges per nodes within hex.
Density	Population density	Number of inhabitants within hex.
Distance to transit	Transit density	Number of transit stops within hex.
Demographics	Social status static	Social status within hex.
	Social status dynamic	Change of social status within hex.
	Income	Average Income within hex.
Diversity	Land use (x8)[a]	Area of given land use category within hex.
Design	Parking area	Parking area within hex.

[a]For the "diversity" dimension, there is one feature per land use class, resulting in eight features for this dimension.

in social status between 31.12.2018 and 31.12.2016 in comparison to the other planning districts of Berlin Senatsverwaltung für Standtentwicklung und Wohnen (2019). We hypothesise that the deviation of social status over time can indicate the level of gentrification of an area, which can effect travel behavior (Matsuyuki et al., 2020). Next to that, we include household income data on the post code level clustered into seven ordered categorical variables (1 to 7) from the commercial provider AXCIOM. The dataset contains 8,173 individual samples and the data was collected throughout the year 2018.

3.1.2. Data cleaning

Data cleaning is applied to reduce noise in the mobility dataset, which may negatively impact predictive power of the model, and select specifically commuting trips within Berlin. We remove all trips with a distance below 2 km, as many of those short trips represent wrong measurements. We make the assumption that within the city boundaries only an insignificant amount of citizens commutes such short distances by car. We additionally filter only trips that (i) are taken on weekdays between 5:00–10:00 AM, many of which are expected to be commuting trips, (ii) are marked as "non commercial" to focus only on private trips, and (iii) that start and end inside of administrative boundaries of Berlin, as some datasets used to compute the 6D's are not available beyond this area. By applying these data cleaning strategies, the sample size reduces to 3,456,746 unique trips. After cleaning, the distribution of the data tends to follow an exponential distribution with $\lambda = 0.143$ (Further statistics after data cleaning are summarised in SI, "2 Methods: Data Cleaning")

3.1.3. Target variable – Vehicle Kilometer Traveled (VKT)

To generate a target variable that accounts for the variation between trip lengths, we calculate the travel distance based on the waypoint input data, divide Berlin into hexagons using the Python toolbox H3 developed by Uber (Brodsky, 2018) (for a detailed introduction to H3 refer to SI, "3.1 H3 Grid") and take the average distance of all trips originating or respectively ending in a hexagon. The model inputs are calculated at the hexagon level, both for predictive features (which represent the 6D's within an hexagon) and the targets (which is the average trip distances starting from or ending at the respective hexagon). Based on the analysis of three different hexagon grid cell sizes, we adopt a hexagon grid cell size with an average area of 0.74 km^2, resulting in 1130 data points at the origin and 1140 at the destination. In comparison to the other two analyzed grid cell sizes, the chosen resolution solves best the trade-off between generating good prediction results, capturing high spatial accuracy and preserving individual privacy. Refer to SI, "6.3 Modeling results for lower and very high resolution" for detailed results.

3.1.4. Feature engineering

To predict VKT, we develop 18 distinct features at the hexagon level, summarised and explained in Table 1. For "destination accessibility" we develop the features distance to city center, distance to local center, street connectivity 1 and 2. We describe "density" via the feature population density, "distance to transit" via the feature transit density and "demographics" via the features social status (provided as a static measurement and its deviation over time). "Diversity" is reflected as area of land use per hexagon and "design" as parking area per hexagon. We also add a feature "noise" to capture to what extend our model is fitted to a random signal (and thereby overfitted). For a detailed description of all features, refer to SI, "3.2 List of all features".

3.2. Machine learning approach

3.2.1. Predictive models

We investigate the predictive power of urban form features using gradient boosted decision trees (Friedman, 2002), as they can detect nonlinear dependencies that are expected to exist between VKT and urban form data (Ding et al., 2018). Recently, Spadon et al. (2019) have shown that for the family of problems this article investigates, boosted trees have the highest prediction accuracy compared to other commonly used machine learning models. Similar studies also have shown that boosted trees have a high robustness against multicollinearity (Ding et al., 2018). We use the machine learning library eXtreme gradient boosting (XGBoost) as it performed best in the analysis of Spadon et al. (2019) and is designed to be computationally most efficient (Chen and Guestrin, 2016). To ensure a general robustness of our prediction, we also test an alternative GBDT implementation from SCIKIT-LEARN library (Pedregosa et al., 2011) (refer to SI, "6.4 Modeling results using sklearn's gradient boosting regressor"). We compare the results with a simple baseline model that takes the median of VKT distribution per hexagon grid cell.

3.2.2. Evaluation metrics and feature importance calculation

We evaluate our predictions using the coefficient of determination (R^2), the mean absolute error (MAE), and root mean squared error ($RMSE$). To calculate individual feature importance, we use SHAP values as proposed by Lundberg et al. (2018). Refer to SI, "5 Methods: Individual Feature Importance", for a detailed explanation of SHAP. Prior GDBT approaches (Wu et al., 2019; Tao et al., 2020; Ding et al., 2018) did not specify which feature importance measures they used and likely applied default measures, which can be inconsistent according to Lundberg et al. (2018). To keep our results comparable with prior research, we also calculate feature importance based on a default importance measure, which can be found in the SI, "7.1 Feature importance based on 'total gain'.

3.2.3. Feature selection

The first step of the modeling consists of a feature selection process to increase explainability by conserving only features that contribute substantially more than a random noise signal to the prediction (Fig. 1, step 3.1). The noise signal is a single number sampled from a standard normal distribution. We use SHAP values to select relevant features for our final prediction (Marcílio and Eler, 2020). We train our model on 100% of the cleaned dataset using all calculated features. In contrast to Marcílio and Eler (2020), we do not choose a predefined subset of n features and instead select only features that contribute at least 0.5% more than the noise feature. This results in a selection of five features for the model used to predict VKT at the origin, and seven at the destination. Refer to SI, "6.1 Feature selection using SHAP", for a detailed description of the results.

3.2.4. Model tuning and validation

The second step of the modeling focuses on determining the collective impact of the urban form features on VKT (Fig. 1, step 3.2). Here, we take only the selected set of features as inputs to the model. To tune the model, we use grid search to find the best hyperparameter values for number of trees and tree complexity, from ranges recommended in previous literature, which is further described in SI, "4.1 Model Training". To prevent overfitting due to spatial autocorrelation between our training and test data (Schratz et al., 2018; Meyer et al., 2019; Milojevic-Dupont et al., 2020), we train the model by applying a 5-fold spatial cross-validation with individual hyperparameter optimisation per fold. To do so, we separate the area of Berlin into five sectors of equal size (resulting in a 80/20 test/train split), all including both inner and outer city areas. One exemplary fold is displayed in SI, "4.2 k-Fold Spatial Cross Validation".

3.2.5. Final feature importance analysis

The last step consists of an analysis of individual feature importance using the best models (Fig. 1, step 3.3). To reflect the results from the spatial cross-validation, we retrain the model on the whole sample using the average hyperparameters across all folds as well as the selected features, and then calculate SHAP values. We acknowledge that the resulting model is based on 20% more training data. Yet, as proposed by Friedman et al. (2001), we can show that the introduced error is insignificant (see SI, "4.3 Error estimation with increasing training sample size") and control for overfitting using the noise feature. To display the individual contribution of each feature to the prediction, we report the relative feature importance calculated via the average SHAP value per feature divided by the sum of average SHAP value changes of all features. In addition, we show a beeswarm plot per feature to highlight which individual feature instances affect the prediction by what margin. To display the spatial scale, we calculate SHAP dependence plots, which show the effect a single feature has on the prediction outcome over its distribution. To display the spatial scale of the feature impact in urban space, we assign SHAP values to the corresponding latitude and longitude coordinates of each hexagon. We utilise the geospatial analysis tool (kepler.gl, 2021) to visualise each SHAP value at its location and further expand the analysis via displaying the interaction with a second feature using the coloring. For a detailed description of SHAP and its benefits over alternative plotting methods, refer to SI, "5.2 Advantages of SHAP values".

4. Results

With our modeling framework as displayed in Fig. 1, we identify various links between urban form features and commute distances by car as well as related CO_2 emissions in Berlin at trip origins and destinations, both on city-wide and street-level resolution. We utilise the framework to carefully control for spatial interdependencies (e.g., inter-feature and auto-correlation effects) as well as to analyze what urban form features affect commute distance, by what margin and at which specific locations. Our results unfold in five steps that display increasing specific contextual information: we report (i) the overall impact of all features, (ii) the relative impact of individual features, (iii) the directionality of each feature impact, (iv) the non-linearities in feature impact and the relation to other features, and (v) spatial patterns of a feature impact in relation to another feature. The first four parts (i-iv) are centered around the impact on car-commute distances, while the last (v) focuses on induced CO_2 emissions.

Table 2
The 6D's are predictive for car commute distances in Berlin. This table displays model results for the prediction of vehicle kilometers traveled by car for commute (VKT) using the 6D's of compact development calculated at trip origins and destinations. Results from the main model using gradient boosted decision trees (GBDT) are compared to a baseline model, which makes its predictions based on the median of the VKT distribution. We report as performance metrics the coefficient of determination (R^2), mean absolute error (MAE) and root mean squared errors ($RMSE$) on the test set, averaged across each fold of a 5-fold out-of-sample prediction using a 80/20 train/test split (Step 3.2 on Fig. 1). We also report the mean and the standard deviation (*std*) of the overall VKT distribution for the hexagon grid resolution as reference points.

	Origin		Destination	
	GBDT	Median	GBDT	Median
R^2	0.60	–	0.39	–
MAE [m]	1074.4	1831.5	1053.5	1400.4
$RMSE$ [m]	1527.7	2619.2	1455.6	2004.1
mean [m]	10027.3	10027.3	9696.2	9696.2
std [m]	2532.0	2532.0	1997.0	1997.0

4.1. Urban form accounts for two thirds of the explained variance in commute distances

We find that in Berlin the 6D's of compact development hold predictive information on car commute distances. At the trip origin they hold slightly more predictive power than at the trip destination, as our best gradient boosting models are able to reproduce 60% and 39% of the variation in VKT in each case (see Table 2).

Relying on cross-validation and a 80/20 train/test split (see Methods, Machine Learning Approach), the model at the trip origin and destination yield respectively a mean absolute error of 1,074 m and 1,054 m, and a root mean square error of 1,528 m and 1,456 m; those are substantially lower than the standard deviation on the test set (2,532.0 m and 1,997.0 m). In both predictions, the gradient boosting model clearly outperforms a simple baseline model that takes the median VKT in the training dataset as the prediction for all data points. Refer to SI, "6.2 Prediction results and chosen model hyperparameters per fold" and Fig. 1 for extensive details on the model training and evaluation procedures.

4.2. Distance to center and subcenters have most impact on VKT

Distances to activity centers and income play a prominent role in predicting car commute distances, as they contribute about 90% of predictive power of the model, both at the trip origin and destination (respectively 92% and 88%). Proximity to centers and subcenters reduce VKT, while low average income corresponds to increased VKT. Presence of industrial and commercial areas at the destination translates into higher trip distances.

To identify relative impact of individual features, we calculate average SHAP value magnitudes, its values quantifying the contribution that each feature brings to the prediction made by the model (see Methods, Machine Learning Approach). We apply feature selection to ensure that the relative importance is distributed only between features whose importance is substantially above that of a random signal (Step 3.1 on Fig. 1). To analyze the directionality of effects (does a specific feature value increase or decrease VKT), we compute individual SHAP values as beeswarm plots for both trip origin and destination (Fig. 2, corresponding to step 3.3 on Fig. 1).

At the trip origin, the "distance to the city center", "distance to local centers" and "income" have the largest impact on VKT, with relative contributions to the model output of 56%, 23.5% and 12.5% respectively, reaching 92% together. The large impact of these features was also found in previous studies (Ding et al., 2018; Hong et al., 2014; Næss et al., 2017; Ewing and Cervero, 2017). Overall, five features – distance to city center and subcenter, average income, land road area and land use commercial and industrial area – have a relative importance above that of a random signal at the trip origin. Thus, we could identify a clear effect on VKT for three of the 6D's – destination accessibility, demographics and diversity. It is worth noting that due to economic (Alonso, 1964) and cultural (Fisherman, 1996) reasons, higher densities are more present at central than at suburban regions. This implies that the distance to the center also has an indirect effect on other D-variables, including distance to subcenters and diversity in addition to its direct effect. The results are also specific for longer trips, above 2 km distance, implying that the inclusion of shorter trips might have favored more micro level features, such as street connectivity.

Our analysis reveals that trip origins closer to the center (blue) correspond to lower VKT than trips further away (red), Fig. 2. A dense distribution of trips that are located close to the center have VKT shorter than average, down to −1.7 km. The long right tail implies that at longer distances to the city center VKT increases disproportionally. At some origins, we observe a very strong increasing effect, of up to +8 km above the sample average. A similar distribution of the effects can be seen for the feature "distance to subcenter". The magnitude of the effect ranges from −0.5 km to +2.3 km. Higher average income (red) corresponds to shorter VKT (down to −0.8 km) than low average income (blue, up to +2.3 km), contrasting with findings of previous studies (Ewing et al., 2015; Ding et al., 2018). While the relative importance of "income" is less than that of "distance to the subcenter", the long right tail implies that "income" has an equally strong increasing effect at certain origins. The relative contribution of "road area" is 4.6% and a low presence of road area has a slight reducing effect on VKT. "Industrial and commercial areas" located at the trip origin have a relative importance of 3.6%; a high presence of industrial or commercial area corresponds to higher VKT (Fig. 2).

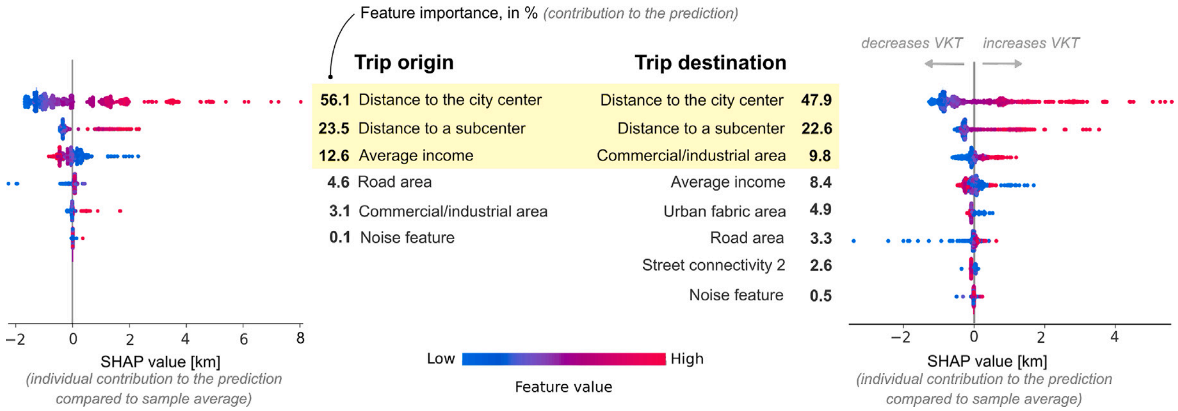

Fig. 2. Distance to activity centers and income have the strongest effect on car commute distances in Berlin. These features account for about 90% of predictive power at trip origins and destinations. For each feature in our model, we report its relative contribution and display how a feature's value (coloring) affects the prediction (x-axis, in kilometer) of vehicle kilometers traveled by car for commute (VKT) for every data point (each dot). If several dots fall onto the same position on the x-axis, i.e. have the same effect magnitude, they accumulate to show density (Lundberg et al., 2020). Three features (highlighted in yellow) capture most of the variation in both models: attraction centers (distance to the main and local centers and commercial / industrial areas, only for destination) have the highest impact, with income also playing a role. VKT increases with distance to center and subcenters. Feature importance (bold values, in %) is calculated as the relative average SHAP value on the training set for model runs using the whole sample and average hyperparameters tuned with spatial cross-validation (Step 3.3 on Fig. 1). The set of beeswarm plots are based on individual SHAP values of each data point of the sample.

At the trip destination, the relative importance is distributed over more features. Seven features – distance to city center and subcenter, average income, land uses commercial/industrial, urban fabric and roads, and street connectivity 2 (defined as the number of streets per intersection) – have a relative importance above that of a random signal, which is two more features than at the trip origin. Our analysis identifies the same three urban form characteristics also at the trip destination — destination accessibility, demographics and diversity. The features "distance to city center" and "distance to a subcenter" still have the largest impact on VKT, with 47.9% and 22.6% respectively, which is 8% and 1% less than in the origin model. Again, we point to the fact that the distance to the city center also has an indirect effect on the other D-variables and that results are specific for trip lengths above 2 km.

Most of the trip destinations close to the city center (blue) correspond to short commute distances, with VKT-reducing effects down to −1.3 km, while for the long tail of trips ending further away (red) there are increasing effects up to +5.8 km, Fig. 2. A similar effect distribution can be seen for "distance to subcenters", with a magnitude from −0.5 km to +3.9 km. "Commercial and industrial area" becomes the third most important feature with a relative importance of 9.8%, while "income" ranks fourth with 8.4%. A high share of industrial and commercial area at the trip destination yields higher predicted VKT (up to +1.8 km), Fig. 2, indicating that such locations must attract commuters from a large catchment area. Trips ending in areas with high average income (red) tend to be shorter than the ones ending in low average income areas (blue), whereas the effect of average income on VKT ranges from −0.5 km up to +1.9 km. The features "urban fabric", "road area" and "street connectivity" have a relative importance 4.9% or less on the model. In particular, the absence of urban fabric slightly increases VKT (up to +0.7 km), the absence of road area strongly reduces VKT (down to −2.9 km) and a high degree of street connectivity slightly reduces VKT (down to −0.1 km).

4.3. Centrality and income operate on distinct spatial scales

Distances to center and subcenter influence car commute distances in highly non-linear manners, permitting the identification of distinct areas that may be targeted to reduce VKT. Both centrality features exhibit a sharp divide between VKT-reducing and VKT-inducing threshold effects when radii of 15 and 6 km respectively are crossed. Low income areas correspond to increased VKT in Berlin, as citizens of such areas might not be able to afford dwellings closer to their workplaces. To illustrate the non-linearities in feature impact and the relation to other features, we report dependence plots of SHAP values for the three features having the largest impact on VKT at the trip origin (Fig. 3). Those display the individual impact on VKT (y-axis, in km) of a feature's values (x-axis) per data point (dot). In addition, we highlight the interaction with a second feature by color coding (Lundberg et al., 2020).

"Distance to city center" exhibits clear thresholds effects (Fig. 3 a.1 and Fig. 3 a.2.). Such effects were previously also observed in Oslo, Norway (Ding et al., 2018). At trip origins located within 5 km to the city center, the feature reduces VKT by −1.7 km, while after 5 km, the increase in distance to city center translates into a strong step-wise increase in predicted VKT. At close to 30 km distance from the city center the effect reaches its maximum of +8 km, indicating that a further spatial expansion of Berlin might result in even longer VKT for citizens living at the outskirts. The coloring unveils patterns in relation to distance to subcenters (Fig. 3 a.1) and income (Fig. 3 a.2). Trip origins where "distance to center" has a negative impact are mostly located very close to a subcenter, while trips with high positive impact tend to be far from any subcenter. Trip origins below the 5-km cut-off are mainly areas of high average income, while areas located further away consist of both low and high income areas and have high to very high SHAP values.

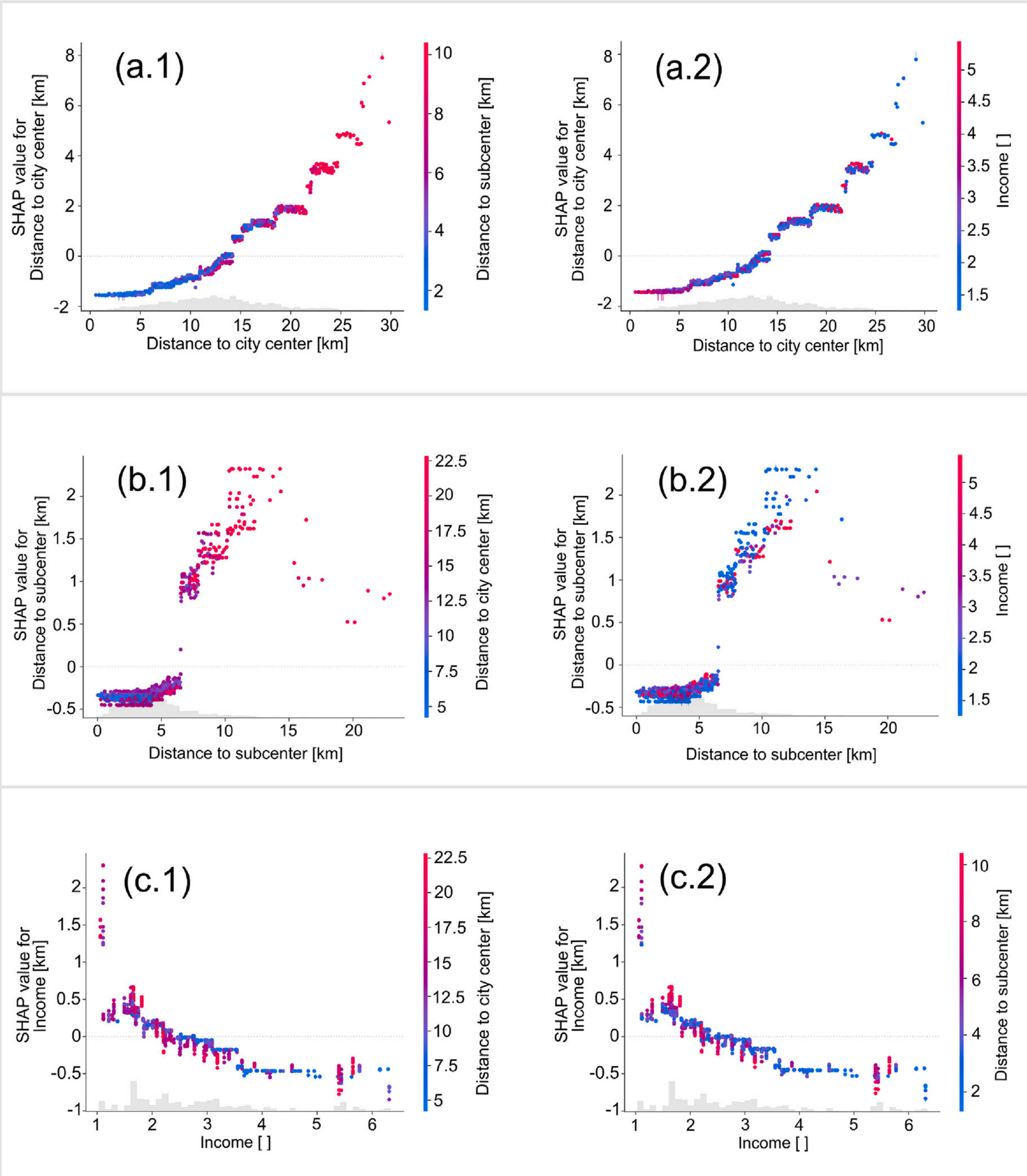

Fig. 3. Further developing subcenters has the potential to reduce car travel in Berlin. Car commute distances rise sharply beyond 10 km from the city center and 6 km from subcenters. In addition, lower average income corresponds to increased commuting distances in Berlin, which motivates to target low-income areas first. SHAP dependence plots for the three most important features at the trip origin showing the non-linearities in individual impact of feature values on predicted vehicle kilometers traveled by car for commute (VKT). Each row corresponds to one feature, displaying the SHAP values (y-axis, left) of each data point (dot) over the feature's marginal distribution (x-axis and visualised via a gray histogram). The two adjacent plots differ in showing a second feature's values via the coloring (y-axis, right). (**a.1**) Distance to city center and distance to subcenters, (**a.2**) Distance to city center and income, (**b.1**) Distance to subcenter and distance to city center, (**b.2**) Distance to subcenter and income, (**c.1**) Income and distance to city center, (**c.2**) Income and distance to subcenter. The SHAP values displayed are computed in Step 3.3 of the modeling framework, see Fig. 1. For the same plots at the trip destination, refer to SI, "8 Results: Spatial scale of feature impact".

More distinct clusters exist between VKT-reducing and VKT-inducing areas for "distance to a subcenter" (Fig. 3 b.1 and Fig. 3 b.2). Most trip origins are clustered within 6 km to the closest subcenter, with a reducing effect on VKT between 0 and −0.5 km.

After that point, an abrupt increase of SHAP values of about +1 km occurs; the increase continues until the distance to subcenter reaches 15 km, forming a second cluster with VKT-inducing effects attaining +2.3 km. The coloring on Fig. 3.C shows that trip origins in the high cluster are also far away from the city center. The low cluster contains diverse distances to the main center and income areas (Fig. 3 b.1 and Fig. 3 b.2). This implies that subcenters have a reducing effect on VKT independent on their location in the city or the average income.

"Income" and VKT are negatively correlated, Fig. 3 c.1 and Fig. 3 c.2. Low-income areas below income class 2 travel longer than average, up to +2 km more in some cases, middle-income areas from class 2 to 4 slightly less than average, and high-income areas above class 4 about consistently less than the average. The coloring in Fig. 3 c.1 shows that lower-income areas tend to live further away from the city center, while other income areas are more evenly distributed. Among middle-income areas, trip origins located within 10 km of the city center (blue coloring) have smaller VKT-reducing effects than the ones located further away (purple and red coloring); a similar pattern exists for distance to subcenters (Fig. 3 c.2). Among low-income areas, the effect of income on VKT is higher for trip origins located beyond 6 km away from a subcenter (purple and red coloring).

4.4. Social equity and sustainability of commutes is linked spatially

Lastly, we find that areas of low-income have disproportionally high commute distances and related increased per capita CO_2 emissions due to their location in the city. By translating the calculated SHAP values of the trip origin model onto a 3D map of Berlin and including location specific CO_2 emission data, the proposed framework allows to analyze the coupled socio-environmental effects of distinct urban form features and display how they alter over space. Each bar represents one location in the city, whereas its height denotes the induced emissions per commute trip. The heights are calculated via multiplying the magnitude of the location's SHAP value with a carbon emission factor, based on the location specific distribution of car types and an average emissions factor per car type, as displayed in SI "9 Results: Calculation of location specific CO_2 emissions". CO_2-reducing areas (in comparison to the sample average) based on negative SHAP values are denoted through a smaller hexagon diameter. The coloring highlights potential interactions with another feature, similar to the dependence plots. We illustrate this map for the feature that impacts VKT most, "distance to city center". The coloring depicts a socio-demographic dimension, "average income" (see Fig. 4.a).

The analysis of the environmental effects reveals that areas closest to the city center correspond to shorter trips and reduced CO_2 emissions. In the outskirts, there is a strong non-linear increase in commute trip related CO_2 emissions in particular towards the south-east (in the areas of Marzahn-Hellersdorf and Treptow Köpenick), and south west (in the areas of Steglitz-Zehlendorf and the lower end of Spandau) of Berlin. When comparing the average of 10% of the highest and lowest CO_2 emissions per area, we find a difference of 0.62 kg CO_2 emissions per trip or 281.5 kg CO_2 emissions per annum (when considering two trips a day and 226 workdays a year). This equals 18.7% of total annual per capita emissions for mobility (including non-commuting trips and excluding air travel) in Germany, which are estimated to be 1.5 tCO_2 (Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit (BMU), 2020).

Fig. 4.b analyzes urban form effects on a socio-demographics dimension by comparing the prevalence of income groups in areas where distance to city center has a negative impact on CO_2 emissions to those with a positive impact (respectively $SHAP-$ and $SHAP+$). In comparison to the $SHAP-$ area, in the $SHAP+$ the share of high income increases by 5%, the share of middle income decreases by 18% and the share of low income increases by 13%. This demonstrates that in Berlin areas of low average income are most affected by disproportionally high VKT and related CO_2 emissions. This reflects strong social inequalities induced by the current urban form characteristics.

5. Discussion

Our findings demonstrate that subcenters have a strong effect on commuting distance both at the trip origin and destination in Berlin. The analysis via dependence plots shows that subcenters at the trip origin reduce VKT slightly higher at the outskirts of the city, where more low-income areas are present. These findings strongly support a more subcenter-oriented development in Berlin in order to reduce car commuting distances. Such developments should especially take place at the outskirts of the city to reduce the high impact of distances to subcenters beyond the threshold of 6 km.

Using their Global Cities Database and a model developed in McIntosh et al. (2014), Newman and Kenworthy (2015) have demonstrated that with increasing density car use declines exponentially as alternative modes become more time effective and people and jobs move closer to each other. While the debate about densification has been most intense in North America, due to the persisting high levels of urban sprawl in cities such as Los Angeles or New York (Barrington-Leigh and Millard-Ball, 2020), it is also highly relevant for Berlin as GHG emissions from the transport sector have been constantly increasing (Senatsverwaltung für Umwelt, 2020).

With our results we can expand on (Newman and Kenworthy, 2015)'s findings and contribute to low carbon urban planning strategies by outlining where the effects of subcenters on VKT unfold. We can show that subcenter-oriented development is not only relevant for reducing car travel via supporting alternative modes. Additionally, subcenters in Berlin have a reducing effect on car driving distance within a 6 km radius. This indicates that policies targeted at realising novel concepts, such as the *15-min city*, also benefit citizens that cannot resign from using their car to commute. Future work can build on our findings and utilise predictive modeling approaches to analyze different development scenarios, similar to the modeling of Silva et al. (2018) applied to the city of Porto. However, with the framework of our study it would not only be possible to model the impact of future development strategies on VKT but also to determine why certain scenarios outperform others and where in particular. By doing so, urban planners and

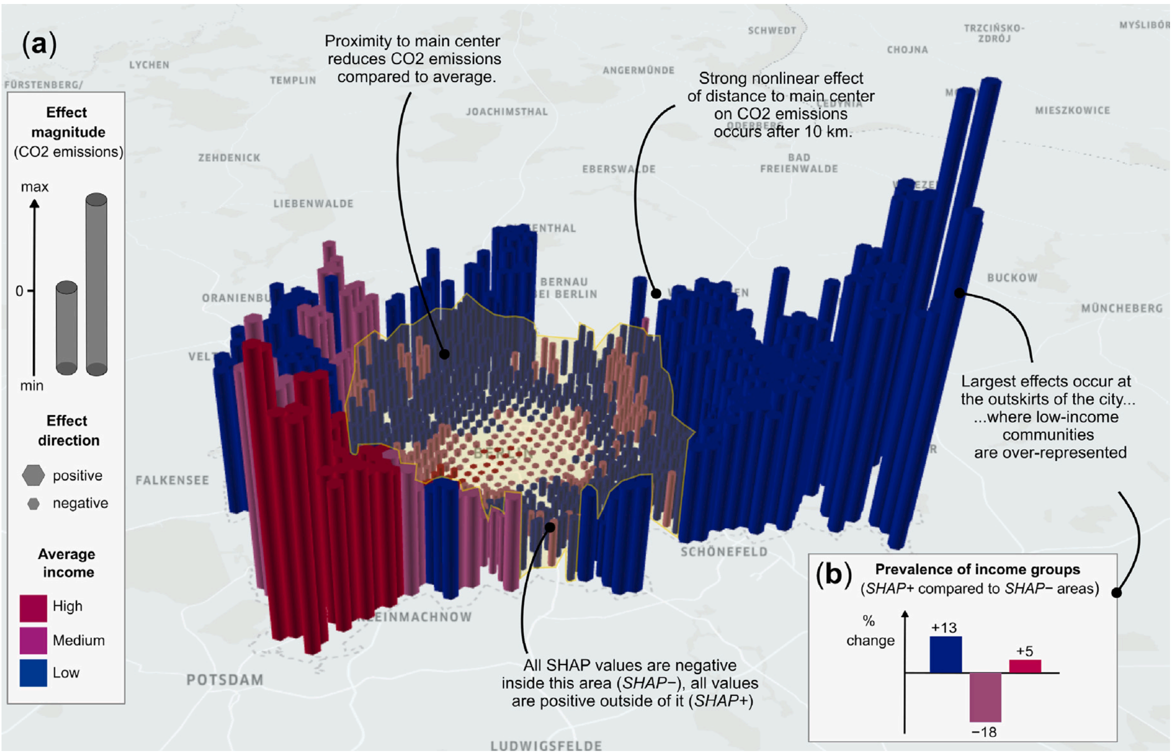

Fig. 4. Identifying specific areas of induced carbon emissions. Commute distances and related CO_2 emissions are strongly increasing towards the south-east and south-west of Berlin, where low-income communities are overrepresented. (**a**) Spatially differentiated influence of the distance to city center on car commute CO_2 emissions (hexagon height) and distribution of income across Berlin (color coding). Proximity to the main center corresponds to reduced CO_2 emissions, whereas areas further away have disproportionally high emissions, especially towards the south-east and south-west of Berlin. Each hexagon represents a data point and its height indicates induced CO_2 emissions based on the SHAP value of the feature "distance to the main center" for a model run using the whole sample and average hyperparameters tuned with spatial cross-validation (Step 3.3 on Fig. 1). SHAP values indicate how much the feature distance to city center value increases the target variable, vehicle kilometers traveled by car for commute (VKT). Negative emissions values (*SHAP-*) are marked by a smaller hexagon diameter and further identified by a light yellow layer, whereas positive ones (*SHAP+*) have a large diameter. (**b**) Difference in prevalence of a given income group between areas with reducing (*SHAP-*) and increasing effect direction (*SHAP+*).

policy makers could be further supported in developing and testing new planning strategies that incorporate a stronger focus on subcenters, their size and their impact on citywide VKT.

We also find that the distance to city center has the largest effect on VKT at the trip origin and destination. The spatial analysis of the effect of distance to the city center at the trip origin reveals areas with disproportionally high CO_2 emissions and strong social inequalities. We find that towards the southeast and southwest of Berlin car-commuting related CO_2 emissions are increasing the strongest and that the distance of residence to the main center can impact total annual per capita footprint for mobility by 18.7%. Considering Berlin's ambitious goals to become climate neutral by 2045 (Senatsverwaltung für Umwelt, 2020), policies to reduce commuting related emissions are required. They should be targeted towards reducing the number of car-commutes in the outer city areas and directing future population growth within a threshold-distance of 15 km to the main center.

Reducing the number of car commuters in the areas with highest induced emissions can be achieved by developing better public transit connections and accessibility (Sims et al., 2014) as well as adopting car sharing and ride hailing initiatives, which focus on increased occupancy per ride (Soukhov and Mohamed, 2022). Next to reduced emissions, less private car use could also free up valuable space in the city that is currently occupied by parking cars (Creutzig et al., 2020). Policies targeted towards densifying the city center should preferably include the development of low income housing to reduce the observed social inequities.

These inequities also have strong implications about where to implement novel urban planning concepts in Berlin, such as the *15-min city*. The preconditions for 15-min neighborhoods (such as density, diversity of residential and commercial areas or already existing subcenters) are currently met predominantly close to the city center. Via the implementation of fast and low-cost interventions, such as opening up bike lanes, micromarkets, pop-up stores or outdoor restaurants, the concept might be realised quickly, without the need of large scale infrastructure development (Moreno et al., 2021). Yet, considering the above-mentioned findings, we argue that changes in urban form in Berlin, including the implementation of 15-min neighborhoods, must require planning strategies that reduce the observed spatial inequities.

Considering that space close to the center is limited, strategies to reduce VKT should also include the development of subcenters in outer areas with low average income — to leverage their reducing effects on VKT within a radius of a 6 km distance. However, previous work has shown that allocating additional industrial and commercial areas towards the outskirts can induce car travel, especially when accommodating highly specialised jobs (in terms of education and experience) with a typically larger catchment area (Wolday et al., 2019; Næss et al., 2019b). If not well designed, such centers can favor car commuting as they are less congested, less accessible via public transport and have more available space for parking compared to the city center. To counter this, restrictions of car accessibility need to accompany the subcenter development, including measures, such as ensuring a high accessibility of the subcenter via public transport (Javaid et al., 2020), having local road pricing schemes (Croci, 2016) or limited car parking opportunities (in the form of low parking space availability and high parking fees) (Christiansen et al., 2017; Wolday et al., 2019).

If more data becomes openly accessible, future research should model the effects of subcenter development at the outskirts and assess whether the findings of Wolday et al. (2019) and Næss et al. (2019b) also hold for Berlin. It should also utilise new data sources to better control for residential self-selection effects via including descriptors of travel attitudes and to better depict the individual 6D variables. In particular, the design variable which is traditionally referred to as the degree of car or pedestrian friendly street design (Stevens, 2017) is only approximated by including the area of parking space (Cervero and Kockelman, 1997). Future research should also model the mobility behavior of different modes and travel patterns to develop well-targeted mobility solutions for the residents living in and outside of 15-min neighborhoods.

6. Conclusion

Our findings demonstrate that analyzing and predicting mobility from urban form with big data leads to rich new insights, highly valuable for low-carbon, sustainable urban development. We find that subcenters play a vital role in reducing VKT for commute in Berlin, which supports the 15-min city hypothesis. Observed threshold effects of induced CO_2 emissions call for policies aiming at inner city densification while releasing peripheral low income communities from car dependence. Considering the potential machine learning has to offer (Rolnick et al., 2019; Milojevic-Dupont and Creutzig, 2020), we argue for more research utilising big data methods to derive spatially explicit urban planning strategies.

Importantly, our results demonstrate the feasibility of a scalable approach to climate change mitigation at street and building level, thus complementing currently prevalent national scale scenarios for climate change mitigation in the IPCC. The ability to visualise the effects of different urban form features graphically could be utilised to interact with citizens throughout the planning process, thereby serving as a basis for transdisciplinary design. To further scale the analysis to other cities, availability of mobility data remains a crucial bottleneck. Our approach based on reducing privacy-sensitive waypoint data to origin–destination-matrices on a hexagon grid showcases what kind of data granularity is relevant for generating context-specific sustainability insights. With more data available, the comparison of spatially explicit urban form effects across different cities becomes possible, thereby creating a new foundation for a big data based urban science supporting global sustainability.

7. Code availability

The code of this work is provided as a freely accessible git repository here: https://github.com/wagnerfe/xml4urbanformanalysis

CRediT authorship contribution statement

Felix Wagner: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Nikola Milojevic-Dupont:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Lukas Franken:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Writing – review & editing. **Aicha Zekar:** Software, Writing – review & editing. **Ben Thies:** Data curation, Writing – review & editing. **Nicolas Koch:** Supervision, Writing – review & editing. **Felix Creutzig:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

For all openly available data used in this study references are provided in chapter 6. The code of this work is provided as a freely accessible git repository, see chapter 7.

Acknowledgments

We thank Nicolai Hans, Ying Jin and Kaveh Jahanshahi for insightful discussions on the manuscript.

This work received funding from the CircEUlar project of the European Union's Horizon Europe research and innovation program under grant agreement 101056810.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.trd.2022.103442.

References

Acuto, M., Parnell, S., Seto, K.C., 2018. Building a global urban science. Nat. Sustain. 1 (1), 2–4.

Alonso, W., 1964. Location and Land Use. Harvard University Press, Cambridge, Mass., United States.

Barrington-Leigh, C., Millard-Ball, A., 2020. Global trends toward urban street-network sprawl. Proc. Natl. Acad. Sci. 117 (4), 1941–1950.

Berlin, G., 2014. Straßenbefahrung 2014. dl-de/by-2-0.

Boarnet, M.G., Bostic, R.W., Rodnyansky, S., Burinskiy, E., Eisenlohr, A., Jamme, H.-T., Santiago-Bartolomei, R., 2020. Do high income households reduce driving more when living near rail transit? Transp. Res. D 80, 102244.

Brodsky, I., 2018. H3: Uber's hexagonal hierarchical spatial index. https://eng.uber.com/h3. (Online; Accessed 07 June 2021).

Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit (BMU), 2020. Klimaschutz in zahlen - fakten, trends und impulse deutscher klimapolitik ausgabe 2020. https://www.bmuv.de/fileadmin/Daten_BMU/Pools/Broschueren/klimaschutz_zahlen_2020_broschuere_bf.pdf. (Online; Accessed 07 March 2021).

Capasso Da Silva, D., King, D.A., Lemar, S., 2020. Accessibility in practice: 20-minute city as a sustainability planning goal. Sustainability 12 (1).

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. Transp. Res. D 2 (3), 199–219.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, Association for Computing Machinery, New York, NY, USA, pp. 785–794.

Choi, K., Paterson, R., 2019. Examining interaction effects among land-use policies to reduce household vehicle travel. J. Transp. Land Use 12 (1), 839–851.

Christiansen, P., Engebretsen, O., Fearnley, N., Usterud Hanssen, J., 2017. Parking facilities and the built environment: Impacts on travel behaviour. Transp. Res. A 95, 198–206.

Copernicus Land Monitoring Service, 2018. Urban atlas 2018. https://land.copernicus.eu/local/urban-atlas. (Online; Accessed 16 August 2021).

Creutzig, F., 2016. Evolving narratives of low-carbon futures in transportation. Transp. Rev. 36 (3), 341–360.

Creutzig, F., Agoston, P., Minx, J.C., Canadell, J.G., Andrew, R.M., Le Quéré, C., Dhakal, S., 2016. Urban infrastructure choices structure climate solutions. Nature Clim. Change 6 (12), 1054–1056.

Creutzig, F., Baiocchi, G., Bierkandt, R., Pichler, P.-P., Seto, K.C., 2015. Global typology of urban energy use and potentials for an urbanization mitigation wedge. Proc. Natl. Acad. Sci. 112 (20), 6283–6288.

Creutzig, F., He, D., 2009. Climate change mitigation and co-benefits of feasible transport demand policies in Beijing. Transp. Res. D 14 (2), 120–131.

Creutzig, F., Javaid, A., Soomauroo, Z., Lohrey, S., Milojevic-Dupont, N., Ramakrishnan, A., Sethi, M., Liu, L., Niamir, L., d'Amour, C.B., Weddige, U., Lenzi, D., Kowarsch, M., Arndt, L., Baumann, L., Betzien, J., Fonkwa, L., Huber, B., Mendez, E., Misiou, A., Pearce, C., Radman, P., Skaloud, P., Zausch, J.M., 2020. Fair street space allocation: Ethical principles and empirical insights. Transp. Rev. 40 (6), 711–733.

Croci, E., 2016. Urban road pricing: A comparative study on the experiences of London, stockholm and Milan. Transp. Res. Proc. 14, 253–262, Transport Research Arena TRA2016.

Ding, C., Cao, X.J., Næss, P., 2018. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in oslo. Transp. Res. A 110, 107–117.

Ding, C., Wang, Y., Xie, B., Liu, C., 2014. Understanding the role of built environment in reducing vehicle miles traveled accounting for spatial heterogeneity. Sustainability 6 (2), 589–601.

Ewing, R., Cervero, R., 2010. Travel and the built environment: A meta-analysis. J. Am. Plan. Assoc. 76 (3), 265–294.

Ewing, R., Cervero, R., 2017. "Does compact development make people drive less?" The answer is yes. J. Am. Plan. Assoc. 83 (1), 19–25.

Ewing, R., Hamidi, S., Tian, G., Proffitt, D., Tonin, S., Fregolent, L., 2018. Testing Newman and Kenworthy's theory of density and automobile dependence. J. Plann. Educ. Res. 38 (2), 167–182.

Ewing, R., Tian, G., Goates, J., Zhang, M., Greenwald, M.J., Joyce, A., Kircher, J., Greene, W., 2015. Varying influences of the built environment on household travel in 15 diverse regions of the United States. Urban Stud. 52 (13), 2330–2348.

Fisherman, R., 1996. Bourgeois Utopias: Visions of Suburbia. In: Readings in Urban Theory. Blackwell, Malden, Mass./Oxford, UK, pp. 23–71.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Statist. Data Anal. 38 (4), 367–378.

Friedman, J., Hastie, T., Tibshirani, R., et al., 2001. The Elements of Statistical Learning, Vol. 1, no. 10. Springer series in statistics New York.

Gehl, J., 2010. Cities for People. Island Press, Washington, United States.

Hanson, S., Hanson, P., 1981. The travel-activity patterns of urban residents: Dimensions and relationships to sociodemographic characteristics. Econ. Geogr. 57 (4), 332–347.

Hong, J., Shen, Q., Zhang, L., 2014. How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. Transportation 41 (3), 419–440.

Jacobs, J., 1961. The Death and Life of Great American Cities. Random House, New York, United States.

Javaid, A., Creutzig, F., Bamberg, S., 2020. Determinants of low-carbon transport mode adoption: Systematic review of reviews. Environ. Res. Lett. 15 (10), 103002.

kepler.gl, 2021. Kepler.gl: Make an impact on your location data. https://kepler.gl. (Online; Accessed 07 June 2021).

Liu, Y., Fang, F., Jing, Y., 2020. How urban land use influences commuting flows in Wuhan, Central China: A mobile phone signaling data perspective. Sustainable Cities Soc. 53, 101914.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.

Lundberg, S.M., Erion, G.G., Lee, S.-I., 2018. Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888.

Maat, K., Timmermans, H.J., 2009. A causal model relating urban form with daily travel distance through activity/travel decisions. Transp. Plan. Technol. 32 (2), 115–134.

Marcílio, W.E., Eler, D.M., 2020. From explanations to feature selection: Assessing SHAP values as feature selection mechanism. In: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images. SIBGRAPI, pp. 340–347.

Matsuyuki, M., Aizu, N., Nakamura, F., Leeruttanawisut, K., 2020. Impact of gentrification on travel behavior in transit-oriented development areas in Bangkok, Thailand. Case Stud. Transp. Policy 8 (4), 1341–1351.

McIntosh, J., Trubka, R., Kenworthy, J., Newman, P., 2014. The role of urban form and transit in city car dependence: Analysis of 26 global cities from 1960 to 2000. Transp. Res. D 33, 95–110.

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications–moving from data reproduction to spatial prediction. Ecol. Model. 411, 108815.

Milojevic-Dupont, N., Creutzig, F., 2020. Machine learning for geographically differentiated climate change mitigation in urban areas. Sustainable Cities Soc. 102526.

Milojevic-Dupont, N., Hans, N., Kaack, L.H., Zumwald, M., Andrieux, F., de Barros Soares, D., Lohrey, S., Pichler, P.-P., Creutzig, F., 2020. Learning from urban form to predict building heights. PLoS One 15 (12), e0242010.
Moreno, C., Allam, Z., Chabaud, D., Gall, C., Pratlong, F., 2021. Introducing the "15-Minute City": Sustainability, resilience and place identity in future post-pandemic cities. Smart Cities 4 (1), 93–111.
Naess, P., 2011. 'New urbanism' or metropolitan-level centralization? A comparison of the influences of metropolitan-level and neighborhood-level urban form characteristics on travel behavior. J. Transp. Land Use 4 (1), 25–44.
Næss, P., Cao, X.J., Strand, A., 2017. Which D's are the important ones? The effects of regional location and density on driving distance in Oslo and Stavanger. J. Transp. Land Use 10 (1).
Næss, P., Stefansdottir, H., Peters, S., Czepkiewicz, M., Heinonen, J., 2021. Residential location and travel in the Reykjavik Capital Region. Sustainability 13 (12).
Næss, P., Strand, A., Wolday, F., Stefansdottir, H., 2019a. Residential location, commuting and non-work travel in two urban areas of different size and with different center structures. Prog. Plan. 128, 1–36, Residential location, commuting and non-work travel in two urban areas of different size and with different center structures.
Næss, P., Tønnesen, A., Wolday, F., 2019b. How and why does intra-metropolitan workplace location affect car commuting? Sustainability 11 (4).
Newman, P., Kenworthy, J., 1989. Cities and Automobil Dependence. An International Source Book. Gower Publishing, Brookfield, VT United States.
Newman, P., Kenworthy, J., 1999. Sustainability and Cities: Overcoming Automobile Dependence. Island Press.
Newman, P., Kenworthy, J., 2015. The End of Automobile Dependence. Island Press, Washington, DC, United States.
OpenStreetMap, 2021. https://openstreetmap.org. (Online; Accessed 16 August 2021).
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Körding, K.P., Gomes, C.P., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J.T., Bengio, Y., 2019. Tackling climate change with machine learning. arXiv, arXiv:1906.05433.
Salon, D., 2015. Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose. Res. Transp. Econ. 52, 34–45.
Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2018. Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data.
Senatsverwaltung für Standtentwicklung und Wohnen, 2019. Monitoring soziale stadtentwicklung. https://stadtentwicklung.berlin.de/planen/basisdaten_stadtentwicklung/monitoring/de/indikatorenblaetter. (Online; Accessed 07 June 2021).
Seto, K.C., Dhakal, S., Bigio, A., Blanco, H., Delgado, G.C., Dewar, D., Huang, L., Inaba, A., Kansal, A., Lwasa, S., et al., 2014. Human Settlements, Infrastructure and Spatial Planning. Cambridge University Press.
Silva, M.C., Horta, I.M., Leal, V., Oliveira, V., 2017. A spatially-explicit methodological framework based on neural networks to assess the effect of urban form on energy demand. Appl. Energy 202, 386–398.
Silva, M., Leal, V., Oliveira, V., Horta, I.M., 2018. A scenario-based approach for assessing the energy performance of urban development pathways. Sustainable Cities Soc. 40, 372–382.
Sims, R., Schaeffer, R., Creutzig, F., Cruz-Núñez, X., D'Agosto, M., Dimitriu, D., Sausen, R., 2014. IPCC Fifth Assessment Report (AR5)-chapter 8: Transport.
Soukhov, A., Mohamed, M., 2022. Occupancy and GHG emissions: Thresholds for disruptive transportation modes and emerging technologies. Transp. Res. D 102, 103127.
Spadon, G., de Carvalho, A.C., Rodrigues-Jr, J.F., Alves, L.G., 2019. Reconstructing commuters network using machine learning and urban indicators. Sci. Rep. 9 (1), 1–13.
Stevens, M.R., 2017. Does compact development make people drive less? J. Am. Plan. Assoc. 83 (1), 7–18.
Tao, T., Wu, X., Cao, J., Fan, Y., Das, K., Ramaswami, A., 2020. Exploring the nonlinear relationship between the built environment and active travel in the twin cities. J. Plann. Educ. Res. 0739456X20915765.
Senatsverwaltung für Umwelt, V.u.K.B., 2020. Monitoring-bericht zur umsetzung des Berliner energieund klimaschutzprogramms (bek 2030) berichtsjahr 2020. https://www.berlin.de/sen/uvk/klimaschutz/klimaschutz-in-der-umsetzung/projekte-monitoring/bek-berichte/. (Online; Accessed 07 June 2021).
van de Coevering, P., Maat, K., van Wee, B., 2021. Causes and effects between attitudes, the built environment and car kilometres: A longitudinal analysis. J. Transp. Geogr. 91, 102982.
Wang, D., Lin, T., 2019. Built environment, travel behavior, and residential self-selection: A study based on panel data from Beijing, China. Transportation 46 (1), 51–74.
Watts, N., Adger, W., Agnolucci, P., Blackstock, J., Byass, P., Cai, W., Chaytor, S., Colbourn, T., Collins, M., Cooper, A., Cox, P., Depledge, J., Drummond, P., Ekins, P., Galaz, V., Grace, D., Graham, H., Grubb, M., Haines, A., Hamilton, I., Hunter, A., Jiang, X., Li, M., Kelman, I., Liang, L., Lott, M., Lowe, R., Luo, Y., Mace, G., Maslin, M., Nilsson, M., Oreszczyn, T., Pye, S., Quinn, T., Svensdotter, M., Venevsky, S., Warner, K., Xu, B., Yang, J., Yin, Y., Yu, C., Zhang, Q., Gong, P., Montgomery, H., Costello, A., 2015. Health and climate change: Policy responses to protect public health. Lancet 386 (10006), 1861–1914.
Weng, M., Ding, N., Li, J., Jin, X., Xiao, H., He, Z., Su, S., 2019. The 15-minute walkable neighborhoods: Measurement, social inequalities and implications for building healthy communities in urban China. J. Transp. Health 13, 259–273.
Wolday, F., Naess, P., Tønnesen, A., 2019. Workplace location, polycentricism, and car commuting. J. Transp. Land Use 12 (1), 785–810.
Woodcock, J., Edwards, P., Tonne, C., Armstrong, B., Ashiru, O., Banister, D., Beevers, S., Z Chalabi, Z.C., Cohen, A., Franco, O., Haines, A., Hickman, R., Lindsay, G., Mittal, I., Mohan, D., Tiwari, G., Woodward, A., Roberts, I., 2009. Public health benefits of strategies to reduce greenhouse-gas emissions: Urban land transport. Lancet 374 (9705), 1930–1943.
Wu, X., Tao, T., Cao, J., Fan, Y., Ramaswami, A., 2019. Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions. Transp. Res. D 75, 1–12.
Zensus, 2011. Bevölkerung im 100 Meter-Gitter. https://zensus2011.de. (Online; Accessed 07 June 2021).
Zhao, P., Bai, Y., 2019. The gap between and determinants of growth in car ownership in urban and rural areas of China: A longitudinal data case study. J. Transp. Geogr. 79, 102487.